

The Detection of Mislabeledled Trees in the International Cocoa Genebank, Trinidad (ICG,T) and Options for a Global Strategy for Identification of Accessions

Olivier Sounigo², Yvette Christopher¹, Frances Bekele¹, Vishnarayan Mooleedhar¹ and Felicia Hosein³

1 - Cocoa Research Unit, University of the West Indies, Saint-Augustine, Trinidad and Tobago

2 - Cirad-Cp, TA 80/02, Avenue Agropolis, 34098 Montpellier Cedex 5, France, Attached to the Cocoa Research Unit

3 - Department of Plant Science, University of the West Indies, Saint-Augustine, Trinidad and Tobago

Abstract

Mislabelling of trees is a common problem for the management of genebanks. A programme was initiated in 1997 in order to evaluate its extent in the International Cocoa Genebank, in Trinidad (ICG,T), using the RAPD technique. The strategy used, the constraints and limitations encountered, and the practical use of the data obtained are described in this presentation. Options for global approaches to identification of cocoa accessions are compared.

Introduction

Misidentification of trees is a common problem in genebanks, whether in cocoa (Figueira *et al.* 1998) or in other species. In order to evaluate the magnitude of this problem in our genebank, a project was initiated in 1997, using the RAPD technique.

Material and methods

Plant material

A total of 546 trees were tested from 132 different accessions (expected to be clones). The numbers of trees compared per accession are indicated in Table 1. The trees were planted in:

- the University Cocoa Research Station (UCRS = genebank, most recently planted,
- at Marper farm (the oldest plantings of many clones), and
- in the fields of the University campus in St. Augustine.

The trees that were analysed were generally chosen in such a way that their pod characteristics were in agreement with what is known about the accession or the population it belongs to.

Methods

The RAPD technique was used, using a protocol adopted at the Cocoa Research Unit (Christopher and Sounigo 1996), on DNA samples extracted according to Johnson *et al.* (1992). Fourteen primers were used, and 39 amplification products were selected, for their intensity and their reproducibility.

Table 1. Number of trees analysed per accession

Number of accessions	Number of trees analysed
13	2
24	3
72	4
13	5
3	6
3	7
1	8
1	10
1	14
1	24

Results and discussion

Magnitude of the variation within accessions

When tree mislabelling problems were found within one accession, the level of genetic difference between the different samples of that accession varied widely. The numbers of primers and markers differentiating trees within accessions with apparently mis-identified trees are indicated in Tables 2 and 3. In a large proportion of cases, the differences observed between two trees within the same accession were due to only one primer and one marker. Differences at the level of only one marker could be due to any of the following factors, in order of decreasing likelihood:

- Reading errors of RAPD bands (since the probability of making only one reading error is much higher than that of making several of them).
- Human errors during the amplification procedure.
- Errors during multiplication of the accessions and/or field planting; if this is this case, the non-matching trees should be genetically rather close to each other.

Table 2. Number of primers differentiating trees within the same accession

Number of primers differentiating trees within the same accession	Number of accessions	%
1	37	40.6
2	16	17.6
3	16	17.6
4	4	4.4
5	9	9.9
6	4	4.4
7	3	3.3
8	1	1.1
10	1	1.1

Where several markers differ between samples of the same accession, this could be due to any of the following:

- Important mistakes during multiplication and/or planting.
- Errors during the DNA extraction (mixing-up of samples).

- Errors during the amplification procedure (if the different markers were generated by a single primer).

In order to determine which of these factors contribute to the differences observed, several verification steps are proposed:

- Verification of reading errors.
- Verification of errors during amplification procedure (redo PCR using the primers showing differences).
- Verification of errors during DNA extraction (re-extract DNA and redo PCR with the primers showing differences).

Table 3. Number of markers differentiating two genotypes within the same accession

Number of markers differentiating two genotypes within the same accession	Number of accessions	%
1	99	39.4
2	47	18.7
3	39	15.5
4	28	11.2
5	16	6.4
6	10	4.0
7	2	0.8
8	5	2.0
9	4	1.6
10	1	0.4

Level of confidence that matching trees are really identical

Where all the trees within an accession were found to be identical using RAPD analysis, the level of confidence that these trees share identical genotypes (X) can be calculated as $1 - P$, where P is the probability that two different trees share the same banding pattern. P was obtained by multiplying the frequencies of all the shared markers. These frequencies were obtained from a diversity study performed on around 400 accessions. The results show that in most cases, this level of confidence was very high (Table 4).

Table 4. Level of confidence (X) that all trees of an accession are truly identical where RAPD analysis can detect no differences within that accession

	Level of confidence			
	X<95%	95%<X<99%	99%<X<99.9%	X>99.9%
Number of Accessions	1	2	7	29
%	2.6	5.1	17.9	74.3

Application of the results to the ICG,T

All the trees analysed will be tagged in a very eye-catching way, according to the following rules. If all the trees appear identical following RAPD analysis, the name of the accession will be written on the tags. In cases where some of the trees of the accession differ:

- The trees which match with the tree from Marper farm will keep the original accession name.
- The trees which do not match with the tree from Marper farm will receive a new accession name (CRU code).
- Where it is not possible to make a comparison with a tree from Marper farm, each of the genotypes within the accession will be assigned a different letter which will be added to the original accession name, for example UF11a, UF11b *etc.*

In cases where most, or all, of the trees analysed were found to differ from the tree from Marper farm, the analysis will be completed for all of the remaining trees of that accession. If this indicates that most or all the trees in the genebank are different to the accession at Marper farm, then the original tree should be used to propagate material for a new plot.

Future of the identity studies at CRU

The verification process will be continued in the ICG,T with priority being given to the commonly used accessions. These include the accessions used as controls in disease resistance studies, those included in the CFC/ICCO/IPGRI Cocoa Germplasm Project core collection and pre-breeding activities, those included in the CAOBISCO Black Pod project for genome mapping and pre-breeding, and those used in flavour testing studies. Due to the high cost and the risk of mistakes generated by the use of the RAPD technique, it seems however useful to find another technique. The use of SSR-PCR markers developed by Lanaud *et al.* (1999) seems very promising to us, since this technique is very discriminating, very reliable and gives genetic information (% of heterozygosity) usable for other types of studies. The use of SCAR obtained from ISSR-PCR markers or other markers might also be appropriate, since the technique has the advantages of extreme simplicity in use and in data management. On the other hand, this technique is probably a little less reliable than the SSR-PCR technique due to the risk of false positives and false negatives (coding as presence/absence of bands). This disadvantage could be reduced, but not eliminated, by the use of appropriate controls.

Global strategy options for the detection of misidentified trees

Similar verification activities should be conducted in all the cocoa genebanks of the world, in order to ensure that researchers using material with the same clone name are actually using material of the same genotype. Four main options are possible for the organisation of such a world-wide characterisation programme, implying different levels of participation and autonomy for the different research centres.

Option 1. The first option requires that all research centres use the same technique and the same markers. With this option, every research centre characterises its accessions, tree by tree, and communicates its individual tree data to the International Cocoa Germplasm Database (ICGD). An example of the data form to be sent by every research centre is given in Table 5.

Table 5. Example of data on mislabelling that could be generated by a research centre and transmitted to the ICGD. M1 to M7 correspond to the markers used. The values 0/1 are indicated if dominant markers are used (presence/absence) and the other values are indicated if a pluri-allelic system is used (genotypes)

Accession	Tree	M1*	M2	M3	M4	M5	M6	M7
UF11	1	1	0	0	0	1	0	0
		2	2	3	6	7	4	1
UF11	2,3,4,5	1	1	1	1	0	0	1
		3	3	3	2	1	8	9

This first option would allow a flexible and continuous process, the only limitations being the communication of the data by the research centres and the storage of these data in the ICGD. This option is however only possible if a technique can be identified which can be used in all the participating laboratories which produces reliable, comparable data which can be easily stored in the ICGD.

Option 2. With the second option, a central laboratory would fingerprint one reference genotype for each accession and communicate the fingerprinting data to all research centres. Each research centre would then fingerprint its own genotypes and compare them to the reference, renaming them if differences are observed. This option would still allow a certain level of flexibility and of autonomy for the research centres, despite the need for a central laboratory, but it implies the need for a common technique and set of markers which could be used to generate data easily comparable between laboratories.

Option 3. The third option requires each research centre to characterise its accessions tree by tree, using the technique of its choice. DNA samples from each of the different genotypes detected in each accession would be sent to a central laboratory, indicating which trees correspond to the different genotypes. A comparison of the samples from the different research centres will be performed in the central laboratory and the data will be sent to these centres and to the ICGD, where they will be stored. An example of the data generated by the central laboratory that can be introduced into the ICGD is shown in Table 6.

Table 6. Example of data generated from the central laboratory on mislabelling of cocoa genotypes in collections

Accession	Country	Tree	Genotype
UF11	Trinidad	1,2,3	a
	Trinidad	5,6	b
	Costa-Rica	1,4	c
	Costa-Rica	2,5	a
	Côte d'Ivoire	1,2	a

This option does not require the use of a universal technique, but requires the use of a central laboratory and is less flexible than option 1. The flexibility could be improved if the procedure used would allow for comparison of data from different experiments, in such a way that there would be no need to wait to have all the samples from one accession available before the analysis could be performed.

Option 4. The fourth option requires each research centre to send DNA samples from each of the trees of all of its accessions to a central laboratory, which would compare all the samples and send the same type of data to the ICGD as in the third option. This option would minimise the amount of work to be done by the different research centres (only DNA extraction) but suffers from a loss of flexibility. Due to the large number of samples to be analysed for each accession, the technique used should imperatively allow the comparison of data from the same accession analysed in different experiments.

Conclusions

The RAPD technique has sufficient discriminatory power to allow us to detect potential labelling mistakes in the ICG,T. The level of variation within accessions was rather high, but could be overestimated by different types of error:

- Errors during the reading of the gels.
- Errors during the amplification procedure.
- Errors during the DNA extraction.

The high percentage of samples from within one accession that differed at the level of only a single marker suggests that reading errors may have been made during the analysis of some samples. In cases where no such errors were made, this indicates that these trees are genetically rather close to each other. Where RAPD analysis detected no variation within an accession, the level of confidence in the similarity of the trees was generally very high (above 99.9% in 74% of the cases and above 99% in 92% of the cases).

It is proposed to rename the accessions according to the results of this study and to replant/establish plots in the genebank with material propagated from the original tree at Marper farm, in cases where too few or no trees were found to be identical to that tree.

We are intending to continue the verification of the ICG,T through the adoption of a less expensive and more reliable technique. The use of PCR-based microsatellites, developed by Lanaud *et al.* (1999) and the use of ISSR-PCR seem to be the most appropriate choices.

Different options have been compared in this paper to establish a global strategy for the detection of mislabelled trees, characterised with different levels of involvement of the research centres in charge of the genebanks.

References

- Figueira A. 1998. Homonymous Genotypes and Misidentification in Germplasm Collections of Brazil and Malaysia. *INGENIC Newsletter* 4: . 4-8.
- Christopher Y. and O. Sounigo. 1996. The use of RAPD for characterization and genetic diversity assessment of cacao. Pages 38-51 *in* Annual Report of the Cocoa Research Unit for 1995, The Cocoa Research Unit, St. Augustine, Trinidad and Tobago.
- Johnson E., J.R. Russel, F. Hosein, W. Powell and R. Waugh. 1992. A laboratory manual for RAPD analyses in cocoa. Internal Cocoa Research Unit Report (unpublished).
- Lanaud C., A.M. Risterucci, I. Pieretti, M. Falque, A. Bouet and P.J.L. LAGODA. 1999. Isolation and characterization of microsatellites in *Theobroma cacao* L. *Mol. Ecol.* 12: 2141-2143.